CODEN: MCEBD4 2 (2) 97–209 (1982)
ISSN: 0270-7306

VOLUME 2 • FEBRUARY 1982 • NUMBER 2

Molecular and Cellular Biology

It is characteristic of Science and Progress that they continually open new fields to our vision—PASTEUR

Published monthly
by the American Society
for Microbiology

High-Efficiency Cloning of Full-Length cDNA

HIROTO OKAYAMA AND PAUL BERG*

Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305

Received 20 October 1981/Accepted 2 December 1981

A widely recognized difficulty of presently used methods for cDNA cloning is obtaining cDNA segments that contain the entire nucleotide sequence of the corresponding mRNA. The cloning procedure described here mitigates this shortcoming. Of the 10^5 plasmid-cDNA recombinants obtained per μg of rabbit reticulocyte mRNA, about 10% contained a complete α- or β-globin mRNA sequence, and at least 30 to 50%, but very likely more, contained the entire globin coding regions. We attribute the high efficiency of cloning full- or nearly full-length cDNA to (i) the fact that the plasmid DNA vector itself serves as the primer for first- and second-strand cDNA synthesis, (ii) the lack of any nuclease treatment of the products, and (iii) the fact that one of the steps in the procedure results in preferential cloning of recombinants with full-length cDNA's over those with truncated cDNA's.

The availability of complementary DNA (cDNA) copies of mRNA's provides an extremely powerful tool for analyzing the structure, organization, and expression of eukaryote genes (4, 5, 9, 10, 16, 19, 36, 38). Aside from the utility of cDNA's for defining the initiation, coding, and termination sequences of mRNA's, their use as hybridization probes makes it possible to search for, isolate, identify, and characterize the corresponding genes from chromosomal DNA. Indeed, it was comparisons between cloned cDNA's and their genomic counterparts that uncovered the existence of intervening sequences (4, 5, 9, 19, 38) and splicing (37) and the occurrence of genomic rearrangements in the formation of functional immunoglobulin genes (3). More recently, the cloning of cDNA copies of RNA virus genomes (e.g., vesicular stomatitis [32], polio [22], and influenza viruses [35]) has opened the way to a more refined understanding of these viruses' structure, replication, and expression, as well as providing a simpler route for the development of antiviral vaccines (17).

With present techniques for cDNA cloning (12, 31, 33, 42) the yield of recombinant DNAs that have full-length cDNA sequences, i.e., cDNA's containing the entire nucleotide sequence of the mature mRNA, is low. Generally, most cDNA clones contain 3'-untranslated and variable amounts of the protein-coding sequence; obtaining cloned cDNA's with complete 5'-untranslated and protein-coding sequences is rarer, particularly if the mRNA codes for a large protein (32, 35). Although such truncated cDNA's are still useful as hybridization probes (15), they cannot direct the synthesis of complete proteins after their introduction into bacterial or mammalian cells via appropriate expression vectors. Consequently, we have sought to devise a cDNA cloning method that increases the probability of obtaining recombinants with full-length cDNA inserts.

Presently, the initial cDNA copy of a mRNA is synthesized with reverse transcriptase (39) using as primer either oligodeoxythymidylate [oligo(dT)] annealed to the polyadenylate [poly(A)] tail (11, 18) or an oligonucleotide annealed to a complementary sequence in the body of the mRNA (36, 40). The quality of the reverse transcriptase, the integrity and secondary structure of the mRNA, and the reaction conditions influence the length of the primary reverse transcript and, therefore, the completeness of the subsequently cloned cDNA. Second-strand synthesis is at best a poorly controlled step, since it relies on the ability of *Escherichia coli* DNA polymerase I (*Pol*I) to use the initial reverse transcript as both a primer and template, the end result being a hairpin double-stranded DNA with the 5' end of the mRNA sequence in the form of a single-stranded loop of variable size and location (11). S1 nuclease digestion of the single-stranded DNA loop, which must precede the addition of homopolymeric tails to the ends of the cDNA, invariably removes portions of the cDNA corresponding to coding or 5'-proximal portions of the mRNA. Several innovations, such as enrichment for the particular mRNA sequence prior to cDNA synthesis (6), fractionation of single-strand or double-strand cDNA to enrich for particular size classes (6, 10, 36), or even alternative procedures for priming the synthesis of the second strand that eliminate the need for the nuclease digestion of 5'-proximal

sequences (24, 33), have improved the yield of full- or nearly full-length cDNA or at least produced cDNA's with intact 5'-proximal nucleotide sequences (24).

Here we describe a modification in the cDNA cloning procedure that permits the recovery, in high yields, of plasmid recombinants with full- or nearly full-length cDNA inserts. The procedure uses a plasmid DNA vector which itself serves as the primer for first- and ultimately second-strand cDNA synthesis; moreover, one of the steps is designed to enrich for recombinants containing full-length cDNA's over those with truncated cDNA's. The procedure has been applied successfully to the cloning of full-length α- and β-globin cDNA's from rabbit reticulocyte mRNA.

MATERIALS AND METHODS

Chemicals, enzymes, and plasmids. Oligo(dT)- and oligodeoxyadenylate [oligo(dA)]-celluloses (both type 6) were purchased from Collaborative Research Inc. The reverse transcriptase was the avian myeloblastosis virus enzyme obtained from J. Beard at the National Institutes of Health. *E. coli* DNA ligase was provided by I. R. Lehman, *E. coli* DNA polymerase I was provided by J. Widom and S. Sherer, and *Eco*RI endonuclease came from J. Carlson (all from Stanford University). Calf thymus terminal deoxynucleotidyl transferase and *E. coli* RNase H were purchased from PL Biochemical Co.; *Hin*dIII and *Kpn*I endonucleases were products of Bethesda Research Laboratories Inc., *Hpa*I, *Ava*II, and *Pvu*II endonucleases were from New England BioLabs, and *Pst*I endonuclease was obtained from Boehringer-Mannheim. The pBR322-SV40 recombinants that provide the vector-primer DNA and linker DNA fragment (Fig. 1) were constructed by S. Subramani. One contained a simian virus 40 (SV40) DNA segment corresponding to map position 0.71 to 0.86 cloned between the *Pvu*II and *Hin*dIII sites of pBR322 DNA, and the other had a segment from map position 0.19 to 0.32 inserted between the *Bam*HI and *Hin*dIII sites of the plasmid DNA.

Preparation of mRNA. Rabbit reticulocyte mRNA, enriched with α-globin mRNA, was prepared from a postpolysomal supernatant of a reticulocyte lysate obtained from phenylhydrazine-treated rabbits (28). The mRNA was recovered after phenol extraction by alcohol precipitation following two cycles of adsorption and elution from an oligo(dT)-cellulose column (1). In recent attempts to prepare cDNA libraries from other cells (H. Okayama and P. Berg, unpublished data) we have found the guanidinium thiocyanate method (7) to be superior for the preparation of mRNA that is readily reverse transcribed.

Preparation of vector primer and oligo dG-tailed linker DNAs. (For a diagram of this procedure, see Fig. 1). A 400-μg sample of pBR322-SV40 (map units 0.71–0.86) DNA was digested at 37°C with 700 U of *Kpn*I endonuclease in a reaction mixture (0.4 ml) containing 6 mM Tris-hydrochloride (pH 7.5), 6 mM $MgCl_2$, 6 mM NaCl, 6 mM 2-mercaptoethanol, and 0.1 mg of bovine serum albumin (BSA) per ml. After 5 h, the digestion was terminated with 40 μl of 0.25 M EDTA (pH 8.0) and 20 μl of 10% sodium dodecyl sulfate (SDS); the DNA was recovered after extraction with water-saturated phenol-$CHCl_3$ (1:1) (hereafter referred to as phenol-$CHCl_3$) and ethanol precipitation. Homopolymer tails averaging 60, but not more than 80, deoxythymidylate (dT) residues per end were added to the *Kpn*I endonuclease-generated termini with calf thymus terminal deoxynucleotidyl transferase as follows. The reaction mixture (0.2 ml) contained 140 mM sodium cacodylate–30 mM Tris-hydrochloride (pH 6.8) as buffer, with 1 mM $CoCl_2$, 0.1 mM dithiothreitol, 0.25 mM dTTP, the *Kpn*I endonuclease-digested DNA, and 400 U of the terminal deoxynucleotidyl transferase. After 30 min at 37°C the reaction was stopped with 20 μl of 0.25 M EDTA (pH 8.0) and 10 μl of 10% SDS, and the DNA was recovered after several extractions with phenol-$CHCl_3$ by ethanol precipitation. The DNA was then digested with 17 U of *Hpa*I endonuclease in 0.2 ml containing 10 mM Tris-hydrochloride (pH 7.4), 10 mM $MgCl_2$, 20 mM KCl, 1 mM dithiothreitol, and 0.1 mg of BSA per ml for 5 h at 37°C. The large DNA fragment, which contained the origin of pBR322 DNA replication and the gene conferring ampicillin resistance, was purified by agarose (1%) gel electrophoresis and recovered from the gel by a modification of the glass powder method (41). The dT-tailed DNA was further purified by adsorption and elution from an oligo(dA)-cellulose column as follows. The DNA was dissolved in 1 ml of 10 mM Tris-hydrochloride (pH 7.3) buffer containing 1 mM EDTA and 1 M NaCl, cooled to 0°C, and applied to an oligo (dA)-cellulose column (0.6 by 2.5 cm) equilibrated

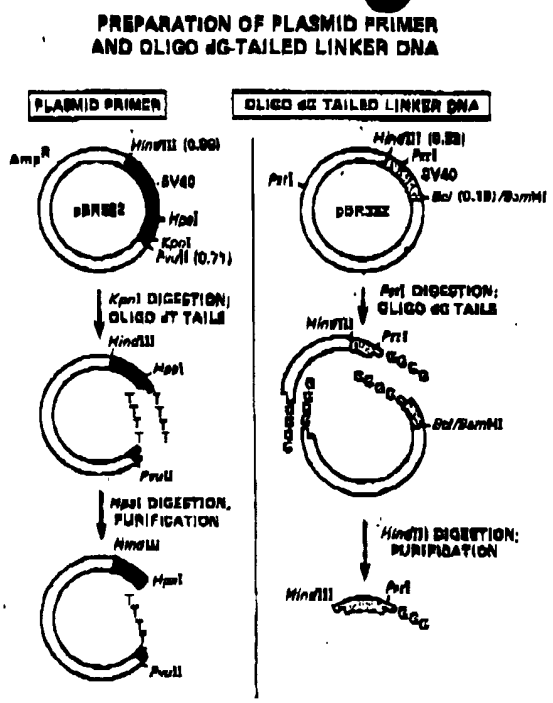

FIG. 1. Preparation of plasmid primer and linker DNA. The unshaded portion of each ring is pBR322 DNA, and the shaded or stippled segments are from SV40 DNA. The numbers next to the restriction site designations are the corresponding SV40 DNA map coordinates.

with the same buffer at 0°C. The column was washed with the same buffer at 0°C and eluted with water at room temperature. The eluted DNA (140 μg) was precipitated with ethanol and dissolved in 100 μl of 10 mM Tris-hydrochloride (pH 7.3) with 1 mM EDTA.

The oligodeoxyguanylate [oligo(dG)]-tailed linker DNA was prepared by digesting 100 μg of pBR322-SV40 (map units 0.19–0.32) with 120 U of *PstI* endonuclease in 0.2 ml containing 6 mM Tris-hydrochloride (pH 7.4), 6 mM $MgCl_2$, 6 mM 2-mercaptoethanol, 50 mM NaCl, and 0.1 mg of BSA per ml. After 1.5 h at 37°C the reaction mixture was extracted with phenol-$CHCl_3$ and the DNA was precipitated with alcohol. Tails of 10 to 15 deoxyguanylate (dG) residues were then added per end with 60 U of terminal deoxynucleotidyl transferase in the same reaction mixture (50 μl) described above, except for 0.1 mM dGTP replacing dTTP. After 20 min at 37°C the mixture was extracted with phenol-$CHCl_3$, and after the DNA was precipitated with ethanol it was digested with 50 U of *HindIII* endonuclease in 50 μl containing 20 mM Tris-hydrochloride (pH 7.4), 7 mM $MgCl_2$, 60 mM NaCl, and 0.1 mg of BSA at 37°C for 1 h. The small oligo (dG)-tailed linker DNA was purified by agarose gel (1.8%) electrophoresis and recovered as described above.

Preparation and cloning of globin cDNA. For a diagram of the preparation and cloning of globin cDNA, see Fig. 2.

Step 1. cDNA synthesis. The reaction mixture (10 μl) contained 50 mM Tris-hydrochloride (pH 8.3), 8 mM $MgCl_2$, 30 mM KCl, 0.3 mM dithiothreitol, 2 mM each dATP, dTTP, dGTP, and [^{32}P]dCTP (850 cpm/pmol), 0.2 μg of the reticulocyte mRNA (1 pmol of globin mRNA), 1.4 μg of the vector-primer DNA (0.7 pmol of primer end), and 5 U of reverse transcriptase. [The molar ratio of poly(A)$^+$ mRNA to vector-primer DNA should be greater than 1.0, and in our experiments ranged from 1.5 to 3). cDNA synthesis was initiated by the addition of reverse transcriptase and continued at 37°C for 20 min. By this time the rate of dCTP incorporation had leveled off and more than 60% of the primer was utilized for cDNA synthesis. The reaction was stopped with 1 μl of 0.25 M EDTA (pH 8.0) and 0.5 μl of 10% SDS; 10 μl of phenol-$CHCl_3$ was added, and the solution was blended vigorously in a Vortex mixer and then centrifuged. After adding 10 μl of 4 M ammonium acetate and 40 μl of ethanol to the aqueous phase, the solution was chilled with dry ice for 15 min, warmed to room temperature with gentle shaking to dissolve unreacted deoxynucleoside triphosphates that had precipitated during chilling, and centrifuged for 10 min in an Eppendorf microfuge. The pellet was dissolved in 10 μl of 10 mM Tris-hydrochloride (pH 7.3) and 1 mM EDTA, mixed with 10 μl of 4 M ammonium acetate, and reprecipitated with 40 μl of ethanol, a procedure which removes more than 99% of unreacted deoxynucleoside triphosphates. The pellet was rinsed with ethanol.

Step 2: Oligodeoxycytidylate [oligo(dC)] addition. The pellet containing the plasmid-cDNA:mRNA was dissolved in 15 μl of 140 mM sodium cacodylate–30 mM Tris-hydrochloride (pH 6.8) buffer containing 1 mM $CoCl_2$, 0.1 mM dithiothreitol, 0.2 μg of poly(A), 66 μM [^{32}P]dCTP (6,000 cpm/pmol), and 18 U of terminal deoxynucleotidyl transferase. The reaction was carried out at 37°C for 5 min to permit the addition of 10 to 15 residues of dCMP per end and then terminated with 1.5 μl of 0.25 M EDTA (pH 8.0) and 0.75 μl of 10% SDS. After extraction with 15 μl of phenol-$CHCl_3$, the aqueous phase was mixed with 15 μl of 4 M ammonium acetate, the DNA was precipitated and reprecipitated with 60 μl of ethanol, and the final pellet was rinsed with ethanol.

Step 3: *HindIII* endonuclease digestion. The pellet was dissolved in 10 μl of buffer containing 20 mM Tris-

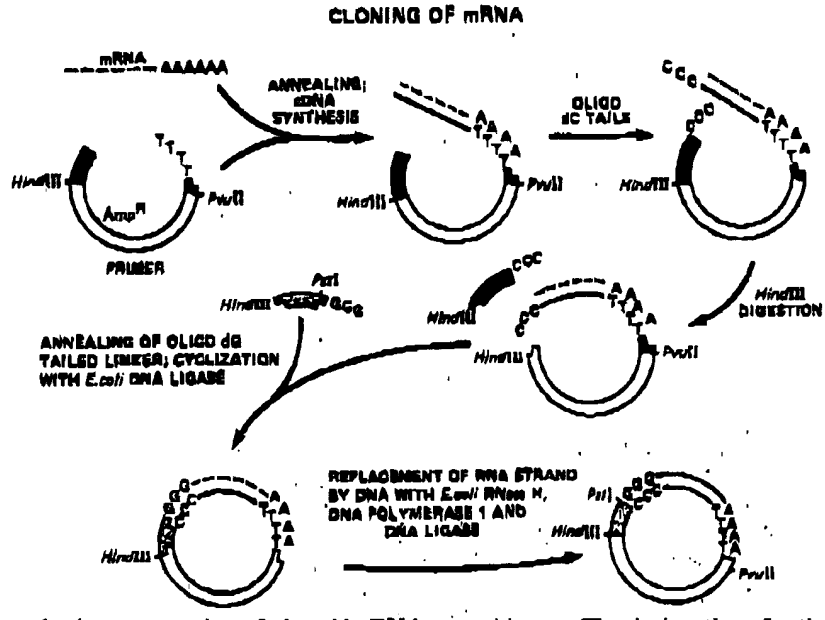

FIG. 2. Steps in the construction of plasmid-cDNA recombinants. The designations for the DNA segments are as mentioned in Fig. 1. Experimental details and comments on the procedure are presented in Materials and Methods and Results, respectively.

hydrochlorides (pH 7.4), 7 mM $MgCl_2$, 60 mM NaCl, and 0.1 mg of BSA per ml and then digested with 2.5 U of *Hin*dIII endonuclease for 1 h at 37°C. The reaction was terminated with 1 μl of 0.25 M EDTA (pH 8.0) and 0.5 μl of 10% SDS, and, after extraction with phenol-$CHCl_3$ followed by the addition of 10 μl of 4 M ammonium acetate, the DNA were precipitated with 40 μl of ethanol. The pellet was rinsed with ethanol and then dissolved in 10 μl of 10 mM Tris-hydrochloride (pH 7.3) and 1 mM EDTA, and 3 μl of ethanol was added to prevent freezing during storage at −20°C.

Step 4: Cyclization mediated by the oligo(dG)-tailed linker DNA. A 1-μl sample of the *Hin*dIII endonuclease-digested oligo(dC)-tailed cDNA:mRNA plasmid (0.02 pmol) was incubated in a mixture (10 μl) containing 10 mM Tris-hydrochloride (pH 7.5), 1 mM EDTA, 0.1 M NaCl, and 0.04 pmol of the oligo(dG)-tailed linker DNA (this amount is a twofold molar excess over the quantity of the vector-cDNA:mRNA and of the fragment which remains as a result of the *Hin*dIII endonuclease digestion in the previous step) at 65°C for 2 min, shifted to 42°C for 30 min, and then cooled to 0°C. The mixture (10 μl) was adjusted to a volume of 100 μl containing 20 mM Tris-hydrochloride (pH 7.5), 4 mM $MgCl_2$, 10 mM $(NH_4)_2SO_4$, 0.1 M KCl, 50 μg of BSA per ml, and 0.1 mM β-NAD; 0.6 μg of *E. coli* DNA ligase was added, and the solution was then incubated overnight at 12°C.

Step 5: Replacement of RNA strand by DNA. To replace the RNA strand of the insert, the ligation mixture was adjusted to contain 40 μM of each of the four deoxynucleoside triphosphates, 0.15 mM β-NAD, 0.4 μg of additional *E. coli* DNA ligase, 0.3 μg of *E. coli* DNA polymerase I, and 1 U of *E. coli* RNase H. This mixture (104 μl) was incubated successively at 12°C and room temperature for 1 h each to promote optimal repair synthesis and nick translation by *Pol*I (23). The reaction was terminated by the addition of 0.9 ml of cold 10 mM Tris-hydrochloride (pH 7.3), and 0.1-ml aliquots were stored at 0°C.

Step 6: Transformation of *E. coli*. Transformation was carried out using minor modifications of the procedure described by Cohen et al. (8). *E. coli* K-12 (strain HB101) was grown to 0.5 absorbancy unit at 600 nm at 37°C in 20 ml of L-broth. The cells were collected by centrifugation, suspended in 10 ml of 10 mM Tris-hydrochloride (pH 7.3) containing 50 mM $CaCl_2$, and centrifuged at 0°C for 5 min. The cells were resuspended in 2 ml of the above buffer and incubated again at 0°C for 5 min; then, 0.2 ml of the cell suspensions was mixed with 0.1 ml of the DNA solution (step 5) and incubated at 0°C for 15 min. After the cells were kept at 37° for 2 min and at room temperature for 10 min, 0.5 ml of L-broth was added, and the culture was incubated at 37°C for 30 min, mixed with 2.5 ml of L-broth soft agar at 42°C, and spread over L-broth agar containing 50 μg of ampicillin per ml. After incubation at 37°C for 12 to 24 h, individual colonies were picked with sterile toothpicks.

Characterization of cDNA clones. (i) Colony hybridization and isolation of recombinant plasmids. *E. coli* transformants were screened for the presence of globin cDNA by in situ colony hybridization (15). One hundred transformants were grown on three replica nitrocellulose filter disks, lysed with alkali, and hybridized with [^{32}P]-cDNA synthesized by oligo(dT)-primed reverse transcription of reticulocyte mRNA. Alternatively, ^{32}P-nick-translated α-globin or β-globin cDNA clones, which had been respectively constructed and identified in the present work or prepared previously (29) from pBG1 (10), were the globin cDNA probes; each of the cDNA's was removed from the vector by restriction enzyme digestions and isolated by agarose gel electrophoresis before use as hybridization probes. Colonies that gave positive hybridization signals were grown in L broth containing 50 μg of ampicillin per ml, and their plasmid DNAs were isolated by standard techniques (21).

(ii) Restriction mapping. DNAs were digested with *Eco*RI, *Pst*I, and *Pvu*II, or *Ava*II restriction endonucleases under conditions recommended by the suppliers, and analyzed by agarose gel (1.5%) electrophoresis.

(iii) DNA sequencing. DNAs were digested with *Pst*I endonuclease, incubated with *E. coli* alkaline phosphatase to remove the terminal phosphates, and terminally labeled with [γ^{32}P]ATP and polynucleotide kinase (26). After digestion with *Eco*RI endonuclease, the ^{32}P-labeled fragment containing the 5′ end of the globin cDNA was purified by agarose gel (1.5%) electrophoresis and sequenced by the method of Maxam and Gilbert (26).

RESULTS

Experimental details of the procedure for preparing and cloning cDNA's are presented above. Here we consider the rationale and several general features of the method and illustrate its application to the cloning of full- and nearly full-length α- and β-globin cDNA's from rabbit reticulocyte mRNA.

Key features of the protocol (outlined in Fig. 2) are that (i) the plasmid vector DNA functions as the primer for the synthesis of the first cDNA strand, an innovation first introduced by Rabbitts (31); (ii) the full- or nearly full-length reverse transcripts of the mRNA are preferentially converted to duplex cDNA's and cloned as recombinants in *E. coli* in the subsequent steps; and (iii) nick-translation repair of the cDNA:mRNA hybrid, mediated by *E. coli* RNase H, *E. coli* *Pol*I, and *E. coli* DNA ligase, is used to synthesize the second cDNA strand.

The vector-primer, a linear DNA with a poly(dT) tail at one end, was constructed as outlined in Fig. 1. The pBR322-SV40 (map μnits 0.71–0.86) DNA recombinant was a convenient precursor because it contains both a unique restriction site (*Kpn*I), at which 3′ single-strand ends can be generated for the efficient attachment of poly(dT) tails, and a second unique restriction site (*Hpa*I) near one end to permit removal of one of the poly(dT) tails. This particular DNA was useful for the purpose described, but other DNAs with similar arrangements of appropriate restriction sites could be substituted.

Annealing of the poly(A)$^+$ mRNA to the poly(dT)-tailed vector-primer DNA generates

the substrate for reverse transcription of the mRNA sequence. A 1.5- to 3-fold molar excess of mRNA over poly(dT)-tailed DNA is advantageous to minimize the possibility that unreacted poly(dT) tails of the vector-primer DNA will be an acceptor for oligo(dC) tails in the next step. Under the conditions used, more than 60% of the vector-primer DNA acquired a covalently linked reverse transcript of the mRNA during the first 10 min of incubation (data not shown); moreover, from pilot experiments with a poly(dT) primer and [^{32}P]dCTP as one of the deoxynucleoside triphosphates, at least 50 to 60% of the cDNA copies attained the length of the globin mRNA.

The aim of the next step is to generate a cohesive tail at the end of the cDNA so that it can be ligated to the other end of the vector DNA and, thereby, provide the template for second-strand cDNA synthesis. This has been achieved by adding oligo(dC) tails to the 3′ ends of both the cDNA and vector DNA and removal of the oligo(dC) tail from the vector DNA terminus by cleavage at the unique *Hin*dIII restriction site near that end. Since the mRNA:cDNA hybrid is a very poor substrate for *Hin*dIII endonuclease (27), it remains intact during digestion with limiting quantities of the restriction enzyme.

The ensuing vector-cDNA:mRNA derivative, with a *Hin*dIII cohesive end and an oligo(dC) tail at the respective termini, is cyclized by *E. coli* DNA ligase using a short linker DNA segment containing a *Hin*dIII cohesive end and an oligo (dG) tail. Covalent joining of the linker and vector DNAs occurs via their *Hin*dIII cohesive ends, and a noncovalent, base-paired join is made to the cDNA:mRNA duplex via the oligo(dG) and oligo(dC) tails. In practice about 20% of the linear vector-cDNA:mRNA was converted to circular structures as judged by the electrophoretic shift in agarose gel.

Since transformation of *E. coli* (HB101) with recombinants containing cDNA:mRNA inserts is inefficient and yields cDNA clones with variable but extensive deletions in the inserts (43), we replaced the mRNA strand by the corresponding DNA strand in vitro. This is accomplished by using *E. coli* RNase H to introduce nicks in the RNA strand (25), *E. coli Pol*I and the four deoxynucleoside triphosphates to replace the RNA segments by nick translation (23), and *E. coli* DNA ligase to join the newly synthesized DNA fragments into a continuous second cDNA strand. In the repair synthesis, the oligo(dG) tail of the linker DNA serves as the primer for copying any unpaired deoxyribosylcytidine (dC) sequence and extending the strand to the cDNA region. Similarly, if necessary, the *E. coli Pol*I extends the oligo(dC) and poly(A) tails to produce complete pairing with the oligo(dG) and poly(dT) of their respective opposite strands. *E. coli* DNA ligase was chosen in place of the T4 enzyme because of its inability to join adjacent RNA and DNA segments (13) that arise during second-strand cDNA synthesis.

Applying this procedure to rabbit reticulocyte mRNA produced about 10^5 ampicillin-resistant HB101 clones per μg of starting mRNA (Table 1). Failure to bridge or ligate the ends of the cDNA and vector by omission of the linker DNA segment or the DNA ligase resulted in drastic reductions in the number of bacterial transformants. If the mRNA strand was not replaced by the second strand of cDNA, the number of bacterial transformants was reduced fivefold.

Characterization of α- and β-globin cDNA clones. To estimate the yield and identify the clones containing α- and β-globin cDNA, colony hybridizations (15) were performed on 100 randomly chosen bacterial transformants (Fig. 3). Figure 3A, B, and C show that about 80% of the transformants contained recombinant plasmids with nucleotide sequences homologous to total reticulocyte cDNA, and 85% of these were accounted for by α-globin (50%) or β-globin (35%) cDNA derivatives. Presumably, the clones with cDNA inserts that failed to hybridize with α- or β-globin cDNA contain nonglobin cDNA's.

A group of clones containing 21 α- or 12 β-globin cDNA's were analyzed further by restriction enzyme digestions to determine the size of the cDNA inserts. The diagram in Fig. 4 summarizes the expected fragments based on the known restriction sites in the recombinant with a complete α-globin cDNA segment. Figure 5 presents the same information for recombinants with a complete β-globin cDNA segment. *Eco*RI endonuclease digestion of the α- and β-globin cDNA recombinant DNAs revealed that 10 of 21 α-globin and 4 of 12 β-globin cDNA clones gave the expected sized fragment (shown by the position of the arrow) for a complete cDNA copy.

TABLE 1. Yield of ampicillin-resistant transformants obtained by cDNA cloning procedure

Procedure[a]	Transformants (no. per μg of mRNA)
Complete	100,000
− Oligo(dG)-tailed linker DNA	2,400
− DNA ligase	50
− DNA polymerase and RNase H	18,000

[a] "Complete" refers to the procedure as described in the text. The other entries refer to modifications in which the indicated component or step was omitted.

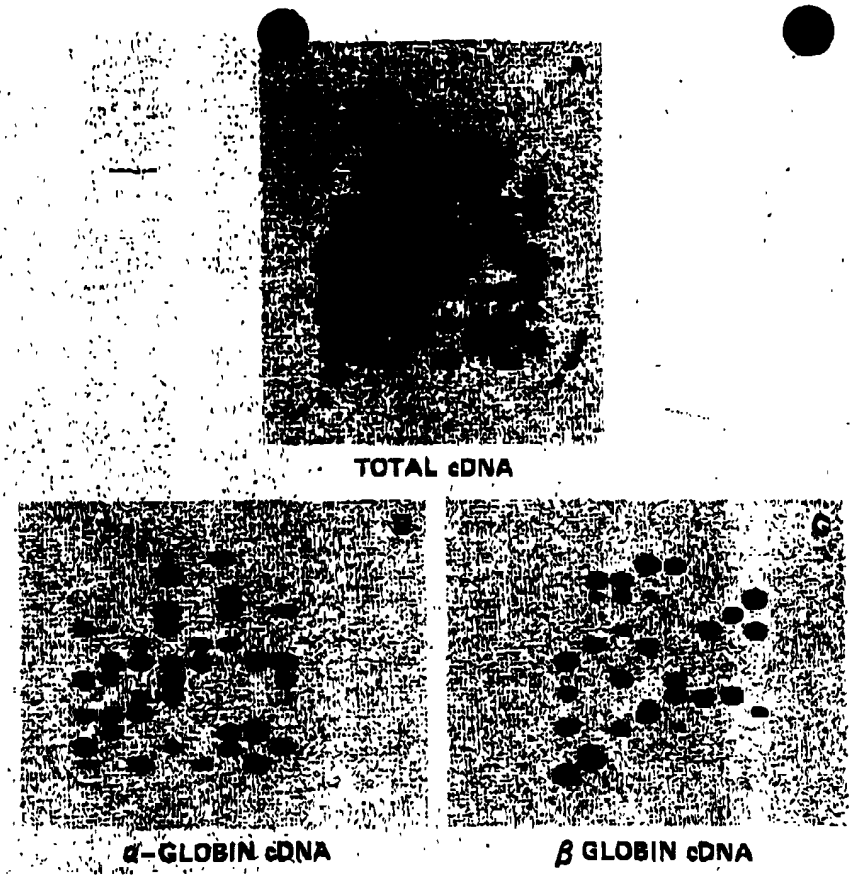

FIG. 3. In situ colony hybridization to detect cDNA clones. One hundred independent ampicillin-resistant transformants were picked and grown on three replica nitrocellulose filters. Detection of cDNA clones was accomplished by colony hybridization (15) using three different probes: (A) total cDNA synthesized by reverse transcription of the rabbit reticulocyte mRNA used for cloning; (B) cloned rabbit α-globin cDNA prepared in this work and identified by restriction analyses; (C) cloned β-globin cDNA made by Efstratiadis et al. (10) and recloned by Mulligan et al. (29).

Each of these plasmid DNAs (10 for α-globin and 4 for β-globin) was digested with *Pst*I endonuclease to cleave at the reconstructed *Pst*I restriction site adjacent to the oligo(dG:dC) join and with *Pvu*II endonuclease to cleave the vector DNA sequence adjacent to the poly(dA:dT) join. All but 2 of the 10 clones, judged to contain nearly full-length α-globin cDNA by the *Eco*RI endonuclease digestion, contained these two restriction sites and yielded the predicted length fragment from a double digest; similarly, 3 of the 4 β-globin cDNA clones, judged to be nearly full length by *Eco*RI endonuclease cleavage, produced the expected sized fragments after *Pst*I-*Pvu*II endonuclease digestions. The fuzzy and heterogeneous bands seen in the *Pst*I and *Pvu*II endonuclease digests of α-globin cDNA clone 2 and β-globin cDNA clone 12 are probably due to heterogeneity in the poly(dA:dT) segment that arises during their propagation in *E. coli*. The eight putative complete α-globin cDNA clones also contain the two known *Ava*II restriction sites, spaced to give the expected sized fragment (Fig. 4).

The 5′-proximal nucleotide sequences of the eight α- and three β-globin cDNA segments were determined to establish the completeness of their cDNA's. After ^{32}P labeling at the reconstructed *Pst*I restriction sites adjacent to the oligo(dG:dC) join, about 120 nucleotides in each cDNA were sequenced by the Maxam-Gilbert method (26) (Fig. 6 and 7). Each figure shows the determined cDNA sequence adjacent to the linker oligo(dG) segment and the known 5′-terminal nucleotide sequence of the corresponding globin mRNA. The nucleotide sequences distal to those shown here were identical to the already reported sequence for rabbit α- (18) and β-globin (10) cDNA's.

Two of the eight α-globin cDNA's contained all the nucleotides at the 5′ end of the mRNA sequence, preceded by 14 or 26 dG residues of the oligo(dG:dC) linker. The other six lacked the first nucleotide at the 5′ end of the mRNA

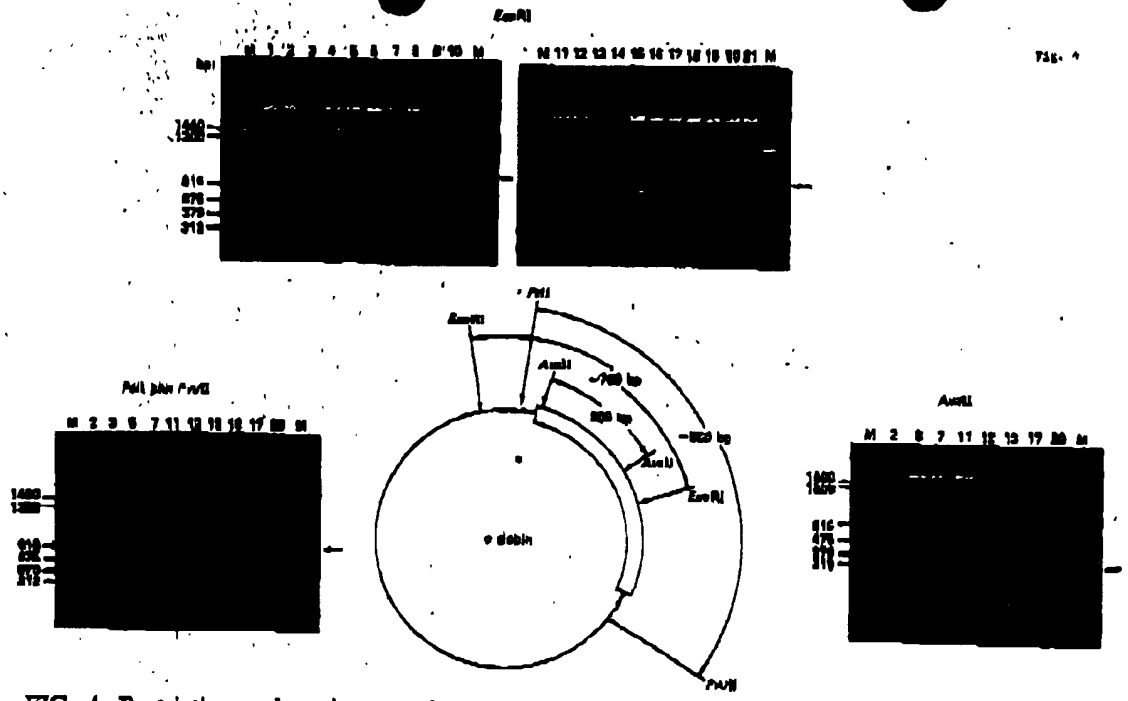

FIG. 4. Restriction endonuclease analyses of α-globin cDNA clones. Plasmid DNA prepared from 21 recombinants with α-globin cDNA inserts were digested with *Eco*RI (upper), *Pst*I and *Pvu*II (left), or *Ava*II (right) endonucleases and electrophoresed on 1.5% agarose gels. The DNA size markers were from *Taq*I endonuclease-digested pBR322 DNA. The presence of a 615-base pair fragment and the relatively low amount of a 475-base pair fragment are due to the presence of a modified and, therefore, resistant *Taq*I restriction site (2). Arrows indicate the position of restriction fragments that would be produced if the cloned cDNA is full length. The diagram in the center summarizes the various enzyme cleavage sites and the sizes predicted for recombinants with complete α-globin cDNA.

sequence and had oligo(dG:dC) linkers ranging in length from 15 to 21 residues. Of the three β-globin cDNA's that were sequenced, one contained the entire 5' end of the mRNA sequence, one lacked one nucleotide, and the other lacked three nucleotides. The linker segments in the β-globin cDNA clones were 14, 16, and 20 nucleotides in length, but the latter two contained, in addition to dG, the unexpected nucleotides dC and dT. These were probably inadvertently introduced because of incomplete removal of dGTP and dATP after the reverse transcription and before the addition of the oligo(dC) tails to the end of the cDNA strand.

Assuming that the cDNA's analyzed are a representative sample of the cloned cDNA population, we estimate that 30 to 50% of the cDNA segments are full- or nearly full-length copies of

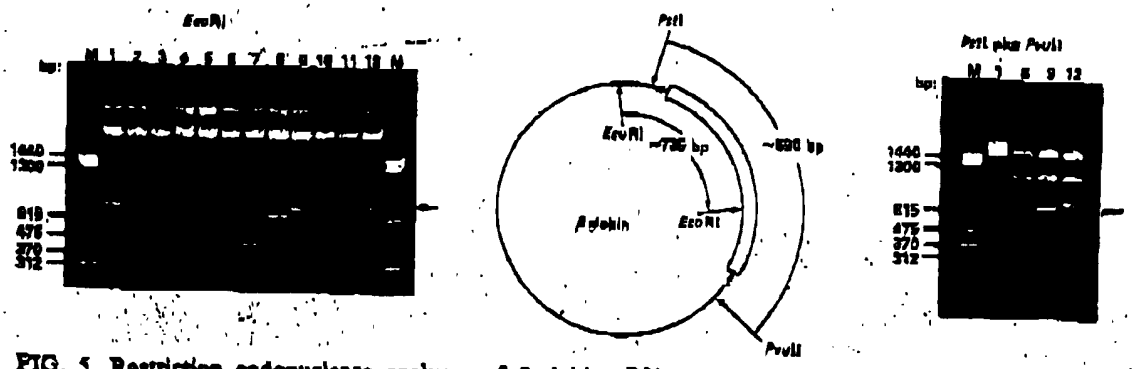

FIG. 5. Restriction endonuclease analyses of β-globin cDNA clones. Plasmid DNA prepared from 12 recombinants with β-globin cDNA inserts were digested with *Eco*RI (left) or *Pst*I and *Pvu*II (right) endonucleases and electrophoresed on 1.5% agarose gels. The markers are the same as in Fig. 4 and the arrows have the same meaning. The diagram summarizes the various enzyme cleavage sites and the sizes predicted for recombinants with complete β-globin cDNA.

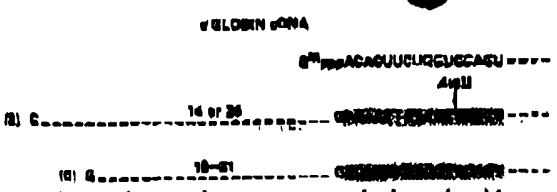

FIG. 6. Comparison of 5'-terminal nucleotide sequences of eight putative full-length α-globin cDNA's with the known 5'-terminal sequence of α-globin mRNA (18).

their mRNA's. Quite likely, still more of the globin clones contain an entire coding sequence since the initiator AUGs for α-globin and β-globin are 36 (18) and 53 (10) nucleotides, respectively, from the 5'-capped nucleotide.

DISCUSSION

The cDNA cloning method described here is general, being applicable to purified or mixed mRNA, to poly(A)$^+$ viral RNAs, or to cellular RNAs to which poly(A) tails can be added at the 3' end. The procedure is relatively rapid, taking less than 2 days from reverse transcription to the collection of transformants containing the cloned cDNA's. Moreover, it uses commercially available reagents and enzymes and standard biochemical procedures. Aside from the relative ease and rapidity, the most significant advantage of this procedure, relative to those previously used, is the high efficiency of cloning full- or nearly full-length cDNA's. Applying the procedure to rabbit reticulocyte mRNA yielded approximately 10^5 globin cDNA clones per μg of mRNA, of which 10% had the complete mRNA sequence at the 5' end and 20 to 40% had all but two or three at the 5'-most nucleotides of the globin mRNA sequences. As anticipated, incorporation of one of the complete α-globin cDNA segments into the vector SVGT5 (29) yielded SV40-like late mRNA's containing the α-globin cDNA sequence and immunoprecipitable α-globin protein (Okayama and Berg, unpublished data).

Although each of the factors contributing to the relatively high cloning efficiency of full-length cDNA's has not been fully evaluated, much of the success, we believe, can be ascribed to the following features. (i) Priming by the plasmid DNA for the synthesis of the first cDNA strand achieves the first join of the cDNA to the vector. (ii) Selective cloning of complete or nearly complete cDNA copies of the mRNA's is necessary because the oligo (dC) tails needed to effect ring closure and complete second-strand cDNA synthesis are synthesized more efficiently by terminal transferase with substrates containing base-paired, full-length reverse transcripts than with shortened or truncated cDNA strands base-paired to longer mRNA's (34); (iii) Replacement of the mRNA by a second cDNA strand is effected by the actions of *E. coli* RNase H (to produce nicks and gaps in the RNA strand), *E. coli Pol*I-mediated nick translation (to replace RNA with DNA sequences), and *E. coli* DNA ligase (to join the DNA fragments into the continuous second cDNA strand).

In this paper the procedure has been used to clone full-length rabbit α- and β-globin cDNA. However, the high cloning efficiency makes it feasible to isolate rare as well as abundant cDNA's from a variety of cellular mRNA populations. For example, a cDNA library containing 2×10^5 clones per μg of mRNA, made with mRNA from Chinese hamster cells, contained several clones that code for the 35,000-dalton subunit of adenine phosphoribosyl transferase, one of which contained a cDNA segment of 0.95 kilobases (kb) and is currently being sequenced to determine its structure (R. Axel, private communication). An analogous cDNA library from hamster cells (A29) that synthesize a methotrexate-resistant dihydrofolate reductase (14) contains cDNA clones whose lengths are sufficient to code for that enzyme (M. Dieckmann, H. Okayama, and P. Berg, unpublished data). In both of these cases the relevant mRNA's are estimated to be present in about 5 to 100 copies per cell. M. Ballivet, J. M. Lee, J. Patrick, and S. F. Heinemann have informed us (private communication) that, using this procedure, they have isolated full- or nearly full-sized cDNA clones that code for the 60,000-dalton subunit of the *Torpedo californica* acetyl choline receptor (20); moreover, Shigesada and Stark (private communication) have obtained a cDNA clone of about 6.5 kb from cell RNA containing the 7.9-kb CAD messenger (30). The efficient production and nucleotide sequencing of nearly full-length cloned cDNA's has also facilitated the characterization of normal and aberrant splicing patterns in their corresponding mRNA's (S. C. Clark, M. C. Nguyen-Huu, H. M. Goodman, and P. Berg, unpublished data). Additional studies to establish the efficiency of this cloning

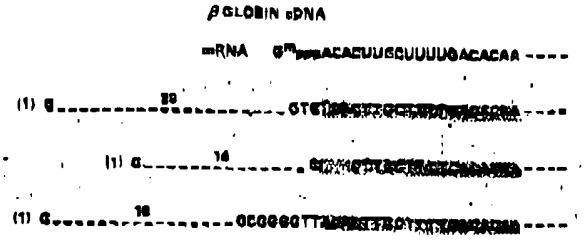

FIG. 7. Comparison of 5'-terminal sequences of three putative full-length β-globin cDNAs with the known 5'-terminal sequence of β-globin mRNA (10).

procedure and the completeness of the cDNA copies of other cellular genes are in progress.

Ours and other approaches for cloning cDNA's (and genomic DNA as well) are clearly limited by the ability to detect particular cDNA's; currently, nucleic acid hybridization (6, 10, 31, 36) or functional assays for the mRNA (16) are the sole means to identify specific cDNA clones. The relatively high yields of cDNA's with a complete protein coding sequence offers another alternative, namely, the identification or even selective propagation of cDNA clones on the basis of the product or function they express. By appropriate choice of the vector-primer and linker DNAs we are exploring the construction of recombinants that can express the cloned cDNA's directly in either bacterial or mammalian cells.

ACKNOWLEDGMENTS

This research was funded by Public Health Service grant GM-13235 from the National Institute of General Medical Sciences, by Public Health Service grant CA-15513-7 from the National Cancer Institute, and by grant MV-35J from the American Cancer Society. H.O. is a fellow of the Jane Coffin Childs Fund for Medical Research.

LITERATURE CITED

1. Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. U.S.A. 69:1408–1412.
2. Backman, K. 1980. A cautionary note on the use of certain restriction endonucleases with methylated substrates. Gene 11:169–171.
3. Brack, C., M. Hirama, R. Lenhard-Schuller, and S. Tonegawa. 1978. Complete immunoglobulin gene is created by somatic recombination. Cell 15:1–14.
4. Breathnach, R., J. L. Mandel, and P. Chambon. 1977. Ovalbumin gene is split in chicken DNA. Nature (London) 270:314–319.
5. Catterall, J. F., J. P. Stein, E. C. Lai, S. L. C. Woo, M. L. Maso, A. R. Means, and B. W. O'Malley. 1979. The ovomucoid gene contains at least six intervening sequences. Nature (London) 278:323–327.
6. Chang, A. C. Y., J. H. Nunberg, R. J. Kaufman, H. A. Erlich, R. T. Schimke, and S. N. Cohen. 1978. Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydrofolate reductase. Nature (London) 275:617–624.
7. Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18:5294–5299.
8. Cohen, S. N., A. C. Y. Chang, and L. Hsu. 1972. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. Proc. Natl. Acad. Sci. U.S.A. 69:2110–2114.
9. Cordell, B., G. Bell, E. Tischer, F. DeNoto, A. Ullrich, R. Dictet, W. J. Rutter, and H. M. Goodman. 1979. Isolation and characterization of a cloned rat insulin gene. Cell 18:533–543.
10. Efstratiadis, A., F. C. Kafatos, and T. Maniatis. 1977. The primary structure of rabbit β-globin mRNA as determined from cloned DNA. Cell 10:571–585.
11. Efstratiadis, A., F. Kafatos, A. M. Maxam, and T. Maniatis. 1976. Enzymatic *in vitro* synthesis of globin genes. Cell 7:279–288.
12. Efstratiadis, A., and L. Villa-Komaroff. 1979. Cloning of double-stranded cDNA, p. 1–14. *In* J. K. Setlow and A. Hollaender, ed., Genetic Engineering, vol. 1. Plenum Press, New York.
13. Fareed, E. G., E. M. Wilt, and C. C. Richardson. 1971. Enzymatic breakage and joining of deoxyribonucleic acid. J. Biol. Chem. 246:925–932.
14. Flintoff, W. F., S. V. Davidson, and L. Siminovitch. 1976. Isolation and partial characterization of three methotrexate-resistant phenotypes from Chinese hamster ovary cells. Somatic Cell Genet. 2:245–261.
15. Grunstein, M., and D. S. Hogness. 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. U.S.A. 72:3961–3965.
16. Harpold, M. M., P. R. Debner, R. M. Evans, and E. C. Bancroft. 1978. Construction and identification by positive hybridization-translation of a bacterial plasmid containing a rat growth hormone structural gene sequence. Nucleic Acids Res. 5:2039–2053.
17. Haseltine, J., and M.-J. Gething. 1981. Cloned copy of the haemagglutinin gene codes for human influenza antigenic determinants in *E. coli*. Nature (London) 292:851–852.
18. Heindell, H. C., A. Liu, G. V. Paddock, G. M. Studnicka, and W. A. Salser. 1978. The primary sequence of rabbit α-globin mRNA. Cell 15:43–54.
19. Jeffreys, A. J., and R. A. Flavell. 1979. The rabbit β globin gene contains a large insert in the coding sequence. Cell 12:1097–1108.
20. Karlan, A. 1980. Molecular properties of nicotinic acetylcholine receptors. Cell Surface Rev. 6:192–246.
21. Katz, L., D. T. Kingsbury, and D. R. Helinski. 1973. Stimulation by cyclic adenosine monophosphate of plasmid deoxyribonucleic acid replication and catabolite repression of the plasmid deoxyribonucleic acid-protein relaxation complex. J. Bacteriol. 114:577–591.
22. Kitamura, N., B. L. Semler, P. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. A. Emini, R. Hanecak, J. J. Lee, S. Van der Werf, C. W. Anderson, and E. Wimmer. 1981. Primary structure, gene organization and polypeptide expression of polio virus RNA. Nature (London) 291:547–553.
23. Kornberg, A. 1980. DNA polymerase 1 of *E. coli*, p. 101–166. *In* DNA replication. H. Freeman and Co., San Francisco.
24. Land, H., M. Grez, H. Hansen, W. Lindemaier, and G. Schnetz. 1981. 5'-Terminal sequences of eucaryotic mRNA can be cloned with high efficiency. Nucleic Acids Res. 9:2251–2266.
25. Leis, J. P., L. Berkower, and J. Hurwitz. 1973. Mechanism of action of ribonuclease H isolated from avian myeloblastosis virus and *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. 70:466–470.
26. Maxam, A., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 65:499–560.
27. Molloy, P. L., and R. H. Symons. 1980. Cleavage of DNA-RNA hybrids by Type II restriction enzymes. Nucleic Acids Res. 8:2939–2946.
28. Morrison, M. R., S. A. Brinkley, J. Gerdil, and J. R. Lingrel. 1974. The separation and identification of α- and β-globin messenger ribonucleic acids. J. Biol. Chem. 249:5290–5295.
29. Mulligan, R. C., B. H. Howard, and P. Berg. 1979. Synthesis of rabbit β globin in cultured monkey kidney cells following infection with a SV40-β globin recombinant genome. Nature (London) 277:108–114.
30. Padgett, R. A., G. M. Wahl, P. F. Coleman, and G. R. Stark. 1979. N-(phosphonacetyl)-L-aspartate-resistant hamster cells overaccumulate a single mRNA coding for the multifunctional protein that catalyzes the first step of UMP synthesis. J. Biol. Chem. 245:974–980.
31. Rabbitts, T. H. 1976. Bacterial cloning of plasmids carrying copies of rabbit globin messenger RNA. Nature (London) 260:221–225.
32. Rose, J. K. 1980. Complete intergenic and flanking gene

sequences from the genome of vesicular stomatitis virus. Cell 19:415–421.
33. Rougeon, F., P. Kourilsky, and B. Mach., 1975. Insertion of a rabbit β-globin gene sequence into an *E. coli* plasmid. Nucleic Acids Res. 2:2365–2378.
34. Roychoudhury, R., E. Jay, and R. Wu. 1976. Terminal labeling and addition of homopolymer tracts to duplex DNA fragments by terminal deoxynucleotidyl transferase. Nucleic Acids Res. 3:863–877.
35. Sleigh, M. J., G. W. Both, and G. G. Brownlee. 1979. The influenza virus haemagglutinin gene: cloning and characterization of a double-stranded DNA copy. Nucleic Acids Res. 7:879–893.
36. Sood, A., D. Pereira, and S. M. Weissman. 1981. Isolation and partial nucleotide sequence of a cDNA clone for human histocompatibility antigen HLA-B by use of an oligonucleotide primer. Proc. Natl. Acad. Sci. U.S.A. 78:616–620.
37. Tilghman, S. M., P. J. Curtis, D. C. Tiemeier, P. Leder, and C. Weissman. 1978. The intervening sequence of a mouse β globin gene is transcribed within the 15S β globin mRNA precursor. Proc. Natl. Acad. Sci. U.S.A. 75:1725–729.
38. Tilghman, S. M., P. J. Curtis, D. C. Tiemeier, J. G. Seidman, B. M. Peterlin, M. Sullivan, J. V. Maizel, and P. Leder. 1978. Intervening sequence of DNA identified in the structural portion of a mouse β globin gene. Proc. Natl. Acad. Sci. U.S.A. 75:1309–1313.
39. Verma, I. M. 1977. The reverse transcriptase. Biochim. Biophys. Acta 473:1–38.
40. Villa-Komaroff, L., A. Efstratiadis, S. Broome, P. Lomedico, R. Tizard, S. P. Naber, W. L. Chick, and W. Gilbert. 1978. A bacterial clone synthesizing proinsulin. Proc. Natl. Acad. Sci. U.S.A. 75:3727–3731.
41. Vogelstein, B., and D. Gillespie. 1979. Preparation and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. U.S.A. 76:615–619.
42. Wood, K. O., and J. C. Lee. 1976. Integration of synthetic globin genes into an *E. coli* plasmid. Nucleic Acids Res. 3:1961–1971.
43. Zain, S., J. Sambrook, R. J. Roberts, W. Keller, M. Fried, and A. R. Dunn. 1979. Nucleotide sequence analysis of the leader segment in a cloned copy of adenovirus 2 fiber mRNA. Cell 16:851–861.